




Check for updates

RESEARCH ARTICLE

Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008-2017 Scopus data [version 1; referees: 3 approved]

Tung Manh Ho^{1,2}, Ha Viet Nguyen², Thu-Trang Vuong³, Quang-Minh Dam²,
Hiep-Hung Pham², Quan-Hoang Vuong ²

¹Institute of Philosophy , Vietnam Academy of Social Sciences, Hanoi , 100000, Vietnam

²Centre for Interdisciplinary Social Research, Western University Hanoi (ĐH Thành Tây), Hanoi, 100000, Vietnam

³Campus Européen de Dijon, Sciences Po Paris, Dijon, 21000, France

v1 First published: 24 Aug 2017, 6:1559 (doi: [10.12688/f1000research.12404.1](https://doi.org/10.12688/f1000research.12404.1))
Latest published: 24 Aug 2017, 6:1559 (doi: [10.12688/f1000research.12404.1](https://doi.org/10.12688/f1000research.12404.1))

Abstract

Background: Collaboration is a common occurrence among Vietnamese scientists; however, insights into Vietnamese scientific collaborations have been scarce. On the other hand, the application of social network analysis in studying science collaboration has gained much attention all over the world. The technique could be employed to explore Vietnam's scientific community.

Methods: This paper employs network theory to explore characteristics of a network of 412 Vietnamese social scientists whose papers can be found indexed in the Scopus database. Two basic network measures, density and clustering coefficient, were taken, and the entire network was studied in comparison with two of its largest components.


Results: The networks connections are very sparse, with a density of only 0.47%, while the clustering coefficient is very high (58.64%). This suggests an inefficient dissemination of information, knowledge, and expertise in the network. Secondly, the disparity in levels of connection among individuals indicates that the network would easily fall apart if a few highly-connected nodes are removed. Finally, the two largest components of the network were found to differ from the entire networks in terms of measures and were both led by the most productive and well-connected researchers.

Conclusions: High clustering and low density seems to be tied to inefficient dissemination of expertise among Vietnamese social scientists, and consequently low scientific output. Also low in robustness, the network shows the potential of an intellectual elite composed of well-connected, productive, and socially significant individuals.

Open Peer Review

Referee Status:   

	Invited Referees		
	1	2	3
version 1			
published 24 Aug 2017	report	report	report

- 1 **Tuyen Quang Tran**, University of Economics and Business, Vietnam
- 2 **Donaldine E . Samson**, Hawaii Pacific University, USA
- 3 **Ly Thi Tran** , Deakin University, Australia

Discuss this article

Comments (2)

Corresponding author: Quan-Hoang Vuong (hoang.vuong@wu.edu.vn)

Author roles: **Ho TM:** Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Nguyen HV:** Data Curation, Formal Analysis, Methodology, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Vuong TT:** Conceptualization, Formal Analysis, Methodology, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Dam QM:** Conceptualization, Data Curation, Investigation, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing; **Pham HH:** Conceptualization, Data Curation, Investigation, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing; **Vuong QH:** Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: The authors declare no conflict of interest.

How to cite this article: Ho TM, Nguyen HV, Vuong TT *et al.* **Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008-2017 Scopus data [version 1; referees: 3 approved]** *F1000Research* 2017, **6**:1559 (doi: [10.12688/f1000research.12404.1](https://doi.org/10.12688/f1000research.12404.1))

Copyright: © 2017 Ho TM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

First published: 24 Aug 2017, **6**:1559 (doi: [10.12688/f1000research.12404.1](https://doi.org/10.12688/f1000research.12404.1))

1. Introduction

In early 2017, the Vietnamese public was once again disappointed to find out there was **no Vietnamese universities** in the Times Higher Education's ranking of the top 300 universities in Asia. There was no shortage of experts' **attempts to explain** this disappointing situation; many pointed to the fact that Vietnamese universities have not put enough focus on research. Being aware of the demand for improving research capacity, the Ministry of Education and Training has recently issued a number of policies and proposals addressing the issue head-on. Figuring among the many efforts is the issuance of **circular No. 08/2017/TT-BGDĐT** (issued on 14th April, 2017) mandating doctoral students must have papers published in Scopus and Web of Science-indexed journals, the doctoral dissertation instructors must also have international publications. There has also been a proposal to mandate that candidates for the titles of Professor and Associate Professor must have international publications. Although these changes and proposals were met with both excitement and dread by the public, it is noteworthy that those who criticize the new regulations do not argue against the changes. Rather, their main concern is "when" or the timeline to adopt these policies: whether these changes are too abrupt.

In other words, people on both sides of the arguments express their desire to improve research capacity in Vietnam. The question remains is "how": How to increase the quantity and quality of scientific publications in Vietnamese social sciences? The answer seems to be related to the spread of information and expertise in the scientific community, which may call for quantitative methods. However, the field of quantitative research on scientific activities and research policy in Vietnam is still nascent. Even though there have been several studies on the status of scientific publications in Vietnam, none has been carried out with a sole focus on social sciences – a field often criticized for having low productivity^{1,2}. In addition, the technique of social network analysis is yet to be applied in the case of Vietnam, despite its potentials in explaining and predicting scientific performance. A study on the nature of scientific co-authorship among Vietnamese social scientists using network statistical analysis would yield valuable insights for policy-makers and educators in Vietnam.

1.1 Literature review

Over the years, the application of network statistical analysis on science collaboration has become pervasive; it has gleaned many insights into the dynamics of scientific activities as well as the properties of scholars' networks. By exploring a number of databases from different fields such as biomedical research, physics and computer science, Newman showed that scientific collaboration networks seem to form "small worlds", in which any randomly chosen pair of scientists would be separated only through a few intermediate collaborators. Another interesting aspect is that there are different degrees of clustering of scientists in different fields, suggesting the differences in social organizations³. In a 2004 study of sociology collaboration networks by exploring of 30 years' worth of data in the field, from 1963 to 1999, Moody discovered that participation in the network depends on the research major, and scholars who are more inclined to quantitative

work are more likely to collaborate than those in non-quantitative work⁴. In 2008, on the relationship between structural and socio-academic communities of co-authorship networks, Rodriguez and Pepe applied different community detection algorithms into the network of scholars in the field of wireless communication and sensors networks. They found out that even in interdisciplinary fields and multi-institutional research groups, co-authorship is heavily influenced by departments and institutional affiliations. In 2010, a study of network analysis on co-authorship and citation networks using topic-modelling path-finding algorithms showed that productive authors tend to cite and directly collaborate with colleagues sharing the same research interests⁵.

Not only the application of network statistics is useful in characterizing the nature of scientist networks, it also provides a powerful tool to study and predict scientific performance such as productivity or research impact. A study on the effects of co-authorship on the performance of scholars using regression model and social network analysis showed that researchers who have a strong connection to only one co-author among a group of connected co-authors perform better than those who have many connections to the same group. The study also suggests it is possible to use professional social network of researchers to predict future performance⁶. In 2013, a group of Taiwanese researchers examined co-authorship networks and research impact through social capital perspective. There are six indicators of social capital in the study: degree centrality, closeness centrality, betweenness centrality, prolific co-author count, team exploration, and publishing tenure. The team found that betweenness centrality is the most influential factor affecting citations of publications⁷. Using data from library and information science in China, a Chinese research team constructed a network of co-authors, then compared an author's centrality values with his/her citations. They found a high correlation between these two elements⁸.

Meanwhile, in Vietnam, network statistics analysis has never been employed to study scientific activities. However, there have been a few attempts to study quantifiable aspects of scientific activities among Vietnamese scholars. Previous studies showed that Vietnam has a low scientific production rate in South East Asia, only equivalent to 13.33% of Singapore and 29% of Thailand in the period of 1991–2010^{7,9}. The total scientific output in Vietnam increased about 16 papers per year during the 1996–2001 period and increased by 20% from 2002 to 2010. It is worth noticing that the share of international collaboration was about 77% of the total publications, of which Japan was the largest collaborating country, followed by the United States, France, South Korea, and United Kingdom^{10,11}. Furthermore, most of the key authors of these international projects did not come from Vietnam but from other countries (Manh 2015)¹⁰. Mathematics was the only field where domestic output proportion was larger than the international. The largest segment was of biology and agriculture, in which 80–90% of published works involved inter-country collaborations. As for social sciences in Vietnam, a study on a sample of 412 Vietnamese scholars who have international publications in Scopus during the period of 2008–2017 has revealed that more than 90% of social scientists have published at least one co-written article (indexed in Scopus), and they worked in collaborations 13 times on average¹².

In short, faced with the current public desire to improve scientific output in social sciences in Vietnam, there is a shortage of in-depth quantitative analysis on the situation of information diffusion and of scientific output in the network of Vietnam scientists. Given the high frequency of co-authoring among social scientists in Vietnam, a network statistical analysis on collaboration among Vietnamese social scientists as the vector of connection would prove to be valuable. It would be interesting to see how network analysis – a technique first developed for studying networks in the natural world – yield valuable insights into the dissemination of knowledge and information of scientific nature among scholars in Vietnam.

1.2. Objectives of the study

This study aims to describe the basic properties of a co-authorship network in a sample of 412 Vietnamese social scientists who have published in Scopus-indexed journals and have online profiles, in the period of 2008–2017.

First, through analyzing the vertex degree distribution in the network, the study will discuss the concept of robustness of the network, which means how well-connected the network could remain if certain nodes and edges are removed. Then through the number of cliques and components, the study will describe the basic structure of the network. Furthermore, using metrics such as density and clustering coefficient, the status of the communication and exchange of scientific knowledge and expertise in the network will be analyzed.

Second, the study does not only provide numerical understanding of the network but also shows various ways in which it can be graphically represented. In doing so, the study will discuss the usefulness of several techniques of network graphical representation that can be applied to facilitate one's understanding of the network.

Finally, the study will extract two of the largest components - one of the largest groups of connected scientists, then explore its characteristics. By comparing this component with the network of 412 Vietnamese social scientists, the study will provide deeper analysis on the concepts visited above.

2. Results

2.1. Characterizing the network of Vietnamese social scientists

Using R, the dataset employed in this paper counts 412 vertices in the Nodes list and 401 edges in the Edges list. Each vertex or node can be different in terms of degree. The average vertex degree is 1.95 with standard deviation 2.26. This means on average, one Vietnamese social scientist co-authors with about two other Vietnamese authors. [Figure 1](#) visualizes the distribution of vertex degrees and shows the disparity between the least and most well-connected authors. ([Figure 1](#) can be plotted using the command in [Supplementary File 1](#) “Rcommands_fig1.doc”.)

An overwhelming majority of researchers - about 280 out of 412 - possesses degree from 0 to 2; only about 50 researchers have a vertex degree of 3–4, and the number of authors with higher degree

decreases dramatically from degree 4 upwards. In other words, most researchers in Vietnam have less than two connections – less than two co-authored papers – and only very few has more than four. Clearly, rather than being composed of mostly people with the same level of connections, the network consists of a few very well-connected people, while the rest does not have many connections at all. It can be inferred that it would be possible to break the network into multiple components if we just removed those few well-connected nodes (people of degree higher than 5) or their links. In network analysis literature, how well-connected a network remains when some vertices and edges are taken out is referred to as robustness¹³. Thus, in this study, the degree distribution reveals that the network of Vietnamese social scientists is not robust. This effect can be seen more visibly when we explore the characteristics of one of the biggest components of this network.

To explore the structure and cohesiveness of the network, it is useful to look at censuses of cliques, components, graph density, and transitivity (Commands for calculation of the network metrics can be found in [Supplementary File 2](#) “Rcommands_metrics.doc”).

By generating a census of cliques of all sizes, we can get a general sense of the structure of the network:

As shown in [Table 1](#), in this network, there are 412 nodes (clique of size 1), 401 edges (clique of size 2), 281 triangles (cliques of size 3), 201 cliques of size 4, and so on. The largest clique is size 9, of which there is only one.

A graph is considered to be connected if every node could be reached by any other node (i.e. if for any two nodes, there is a walk between the two). Looking at [Table 2](#), we can see that the network of Vietnamese social scientists is not connected; there are 125 components of size 1. About 30% of the scientists in this study are isolated nodes in the network, possibly because they either work alone or work exclusively with foreigners. Alternatively, the five biggest components (size 11, 15, 16, 27 and 43) together takes up another 30%, while the rest consists of all middle-sized components (size 2–9).

By calculating the density and transitivity of the graph, it can be seen that the network is very sparse. The density of the graph is 0.0047, indicating only about 0.47% of potential edges are realized in this network. On the other hand, when three vertices are connected at all, there is a better than a 50/50 chance they will form a triangle (clique of size 3): The global clustering coefficient of the collaboration graph is 0.5862, indicating that nearly 59% of connected triples have formed triangles. Given that there is a clear relationship between the speed of the spread of information and clustering coefficient; the higher the clustering coefficient, the slower the information spread¹⁴, it is reasonable to assume when two scientists co-author in a scientific paper, there is a great deal of knowledge and expertise to be communicated and exchanged. Hence, the low density and high clustering coefficient of the network suggests that the dissemination of knowledge and expertise among 412 Vietnamese social scientists in this study is not happening as smoothly as possible.

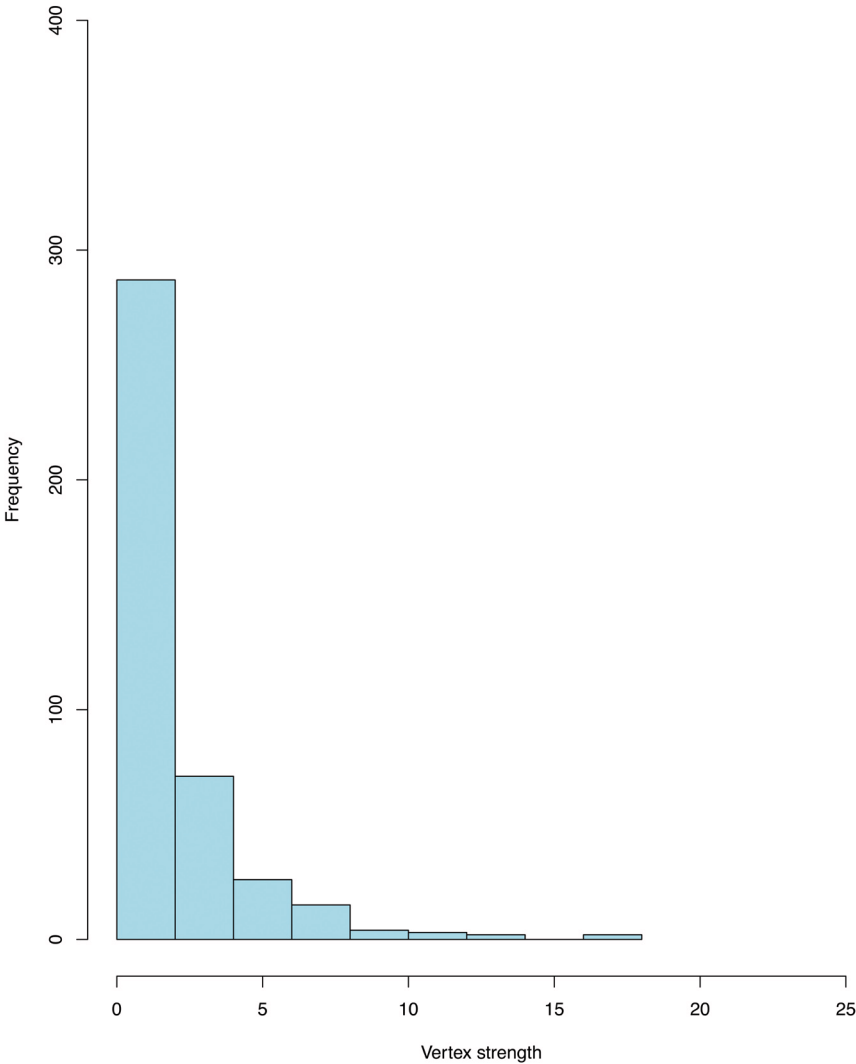


Figure 1. A histogram of vertex degree distribution. This is a histogram of vertex degree distribution of the full 412-node network. The degree of each node is measured as the number of co-authored papers, or connections, of each individual in the dataset.

Table 1. A census of cliques of all sizes for the network of Vietnamese social scientists. A clique is a subset of vertices that are fully cohesive, meaning that all vertices are connected by one link. This table lists all clique sizes that exist within the dataset and the number of cliques in each size category.

Clique size	1	2	3	4	5	6	7	8	9
Numbers	412	401	281	201	144	86	36	9	1

Table 2. A census of components of all size for the network of Vietnamese social scientists. A component is a subgraph in which every vertex can be reached from every other, no matter how many links constitute the path. This table lists all component sizes that exist within the dataset and the number of components in each size category

Component size	1	2	3	4	5	6	7	9	10	11	15	16	27	43
Numbers	125	24	9	3	4	4	2	1	1	2	1	1	1	1

2.2. Network visualization

Visual representations of the network is done through figure plotting in R. Commands for data set-up required for figure plotting can be found in [Supplementary file 3](#) “Rcommands_graph.doc”.

There are several ways to visually represent the network. Here, the study aims to strike a balance between creating a graph both visually attractive and useful in facilitating the statistical understanding of the previous histogram and analysis.

[Figure 2](#) was conceived as a primary representation of the network, highlighting vertex degree, density, transitivity, and robustness using various visual cues. Among many attributes of the nodes that have been collected (region, age, title, etc.), biological gender has been chosen as the basis because of its relatively simple binary nature. In this study, blue color represents male and red represents female. Such simplicity is hoped to make the graph more aesthetically appealing. Meanwhile, the size of each vertex is determined

by the number of edges incident on each node – in other words, by the vertex degree. Hence, the higher the number of edges incident upon a vertex, the bigger the vertex is. This is to make visible the gap between the well-connected scientists and the more isolated ones, one of the most striking features of the network as shown in section 2.1. For layouts, among all those available in R(v3.1.1)’s igraph packages, layout Fruchterman-Reingold is chosen because it makes the structure of the network nicely perceptible: 30% of nodes fall into five largest components, 40% are middle-size components, and the 125 left are isolated nodes (recall the statistics on components in section 2.1). Commands for plotting this figure can be found in the [Supplementary File 4](#) (“Rcommands_fig2.doc”).

Seeking more insights on the network, a community detection algorithm was run on the data, which resulted in [Figure 3](#), a second visualization that complemented [Figure 2](#). (This can be performed using the commands provided in [Supplementary File 5](#) “Rcommands_fig3.doc”). Looking at the biggest components in

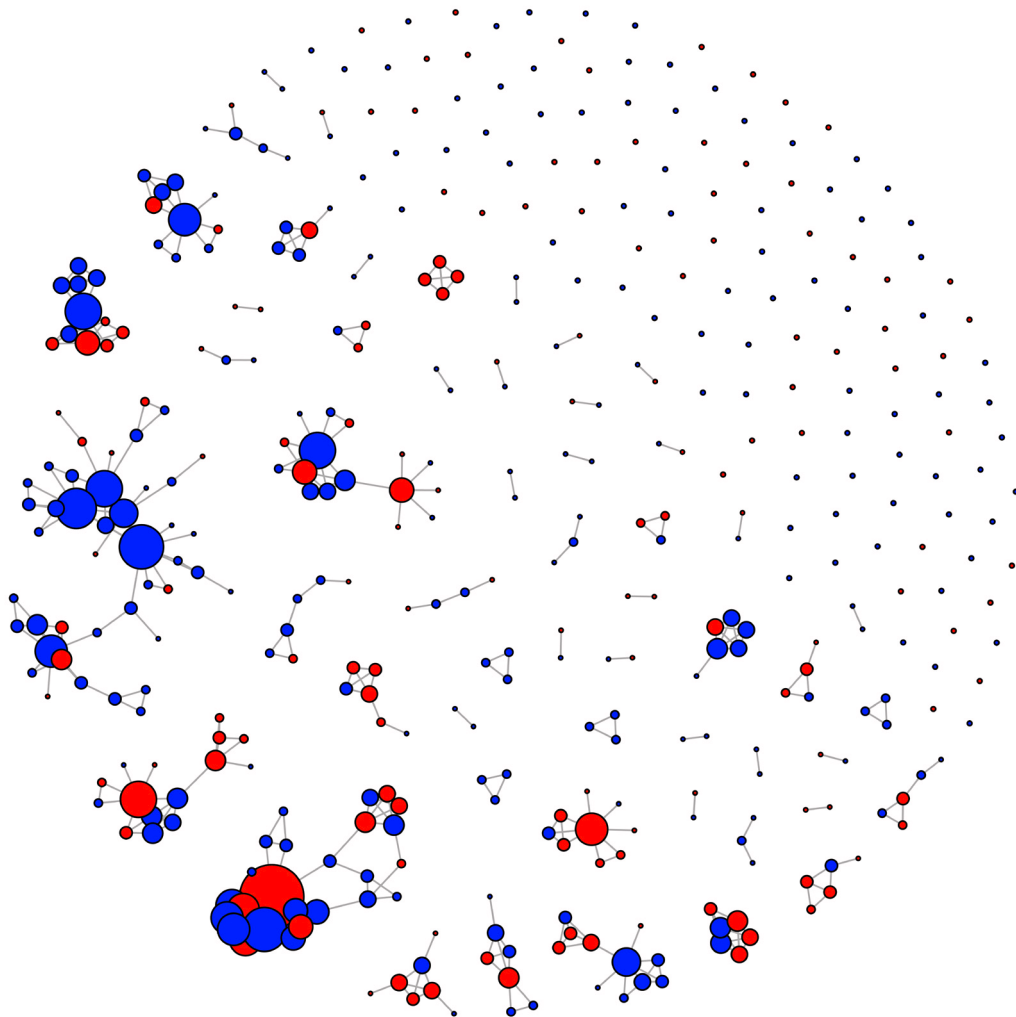


Figure 2. A visual representation of the network of 412 Vietnamese social scientist. This figure is a visualization of the full 412-node network in Fruchterman-Reingold layout. Nodes are color-coded based on author gender (blue for male, red for female). Node sizes are based on node degrees. Edges are represented by a line connecting concerned nodes.

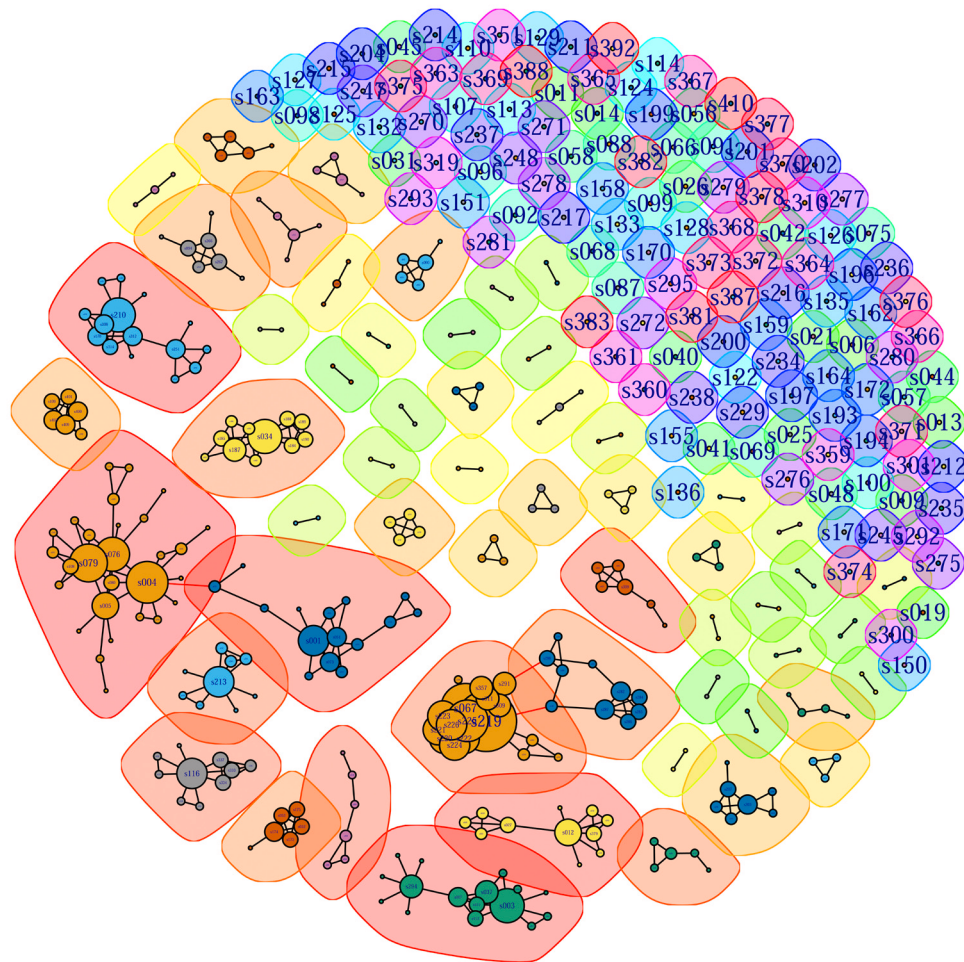


Figure 3. Network visualization with community detection algorithm. This figure is a visualization of the full 412-node network in Fruchterman-Reingold layout with community detection. Potential communities are partitioned using colored regions with boundaries. Colors are mostly to facilitate visual perception and irrelevant to the understanding of the data.

Figure 3, one can see a new pattern emerges: though the big components are fully connected, they do not seem to be one big close group; rather, they seem to consist of a few smaller communities of very closely connected scientists, and these communities are linked together by one or two vertices acting as weak links. The algorithm does indeed break the two big groups into smaller communities with one or two vertices that connect these communities.

In the next section, the two largest components, component size 43 and component size 27, will be studied more in-depth.

2.3. Exploring the characteristics of the two largest components

Recall that component is a technical term in network theory that refers to a maximally connected subgraph, in which any two vertices can be reached from another via a path consisting of any number of edges and nodes. Thus, any graph can be constituted by many different components. In this study, the network of 412 Vietnamese social scientists is the sum total of 179 components of various size, ranging from 1 to 43; the two largest components

have 43 and 27 nodes each. One can treat such components as independent networks in and of themselves. In this section, the characteristics of these two largest components will be explored and compared with the whole network. From this point on, the components will be called Comp43 and Comp27, and the original network will be dubbed Net412. As one might expect, as we zoom in, there will be differences in the properties of the components in question and that of the network as a whole. Table 3 summarizes and compares the basic metrics of Comp43, Comp27 and Net412.

In all network metrics, Comp27 scores the highest. Specifically, in terms of density of connections, Net412 is the sparsest, 0.47%. The density of Comp43 (7.20%) is 14-fold that of Net 412, and the same characteristic in Comp27 (22.51%) is 44-fold compared to that of the whole network. Regarding average vertex degree, Comp27 is the highest followed by Comp43 then Net412. Concerning global clustering coefficient (or transitivity), Comp27 towers over Net412 by 11 percentage points (70% versus 59%), while the latter is in turn over 2 times higher than Comp43 (70% versus 32%).

High clustering and low density suggest a certain level of inefficiency in the spread of knowledge and expertise (as explained in section 1.1 on the characteristics of the network of 412 Vietnamese social scientists); either could be the cause of the other. Thus, from the network metrics, one would expect Comp27's dissemination of scientific knowledge and expertise to be less efficient than Comp43. In fact, even though the density of connection in Comp27 is about 3 times that of Comp43, its effects would be limited because of the higher clustering. One can then ask how to verify that high clustering cancels the good effects of even high density. Supposing that better dissemination of

scientific knowledge and expertise can be observed in a better scientific output, we could look at the mean value of total publications of scientists in each network for insights on the aforementioned question. Indeed, as Table 3 shows, Comp43 performs better than Comp27 in terms of scientific output – almost 3 times higher, 5.53 versus 2.00.

The difference in scientific output between Comp43 and Comp27 can be viewed in Figure 4 below. Commands for plotting Figure 4 (left and right) can be found in Supplementary File 6 and Supplementary File 7 (“Rcommands_fig4left.doc” and “Rcommands_fig4right.doc” respectively.)

Besides revealing the differences in scientific output of the two networks, Figure 4 also reveals that nodes in both networks seems to revolve around one or two important nodes with higher level of scientific output. In Comp43, it is node s004 and in Comp27, it is node s067 and s219 (the visible blue and red dots on the left side of Figure 4). It is interesting that these three nodes have highest numbers of edges incident upon them in their respective networks; s004 has a degree of 11, highest in Comp43; s067 has a degree of 13 and s319 has a degree of 16, also highest in Comp27. If these important vertices are to be removed, the networks would break apart into several smaller components. This feature was referred to in section 5.1 through the concept of robustness, and it should be noted that Net412 is not robust. The situation is the same for Comp43 and Comp27. In Figure 5, the histogram distributing the degrees of nodes in these networks shows a clear disparity in vertex degree.

Table 3. Comparison of basic network metrics of Net412, Comp43 and Comp27. Vertex degree is the number of edges incident upon a vertex. Density is the frequency of realized edges (connections) relative to potential edges (connections). Transitivity (or clustering coefficient) is the relative frequency with which connected triples of vertices form triangles. Net412 is the full 412-node network consisting of the entire dataset. Comp43 and Comp27 are the 43-node and 27-node components, respectively, which are subsets in which every vertex can be reached by every other.

Metrics	Net412	Comp43	Comp27
Graph density	0.47%	7.20%	22.51%
Mean degree	1.95	3.02	5.58
Transitivity	58.62%	32.43%	70.43%
Mean total publications	3.56	5.53	2.00

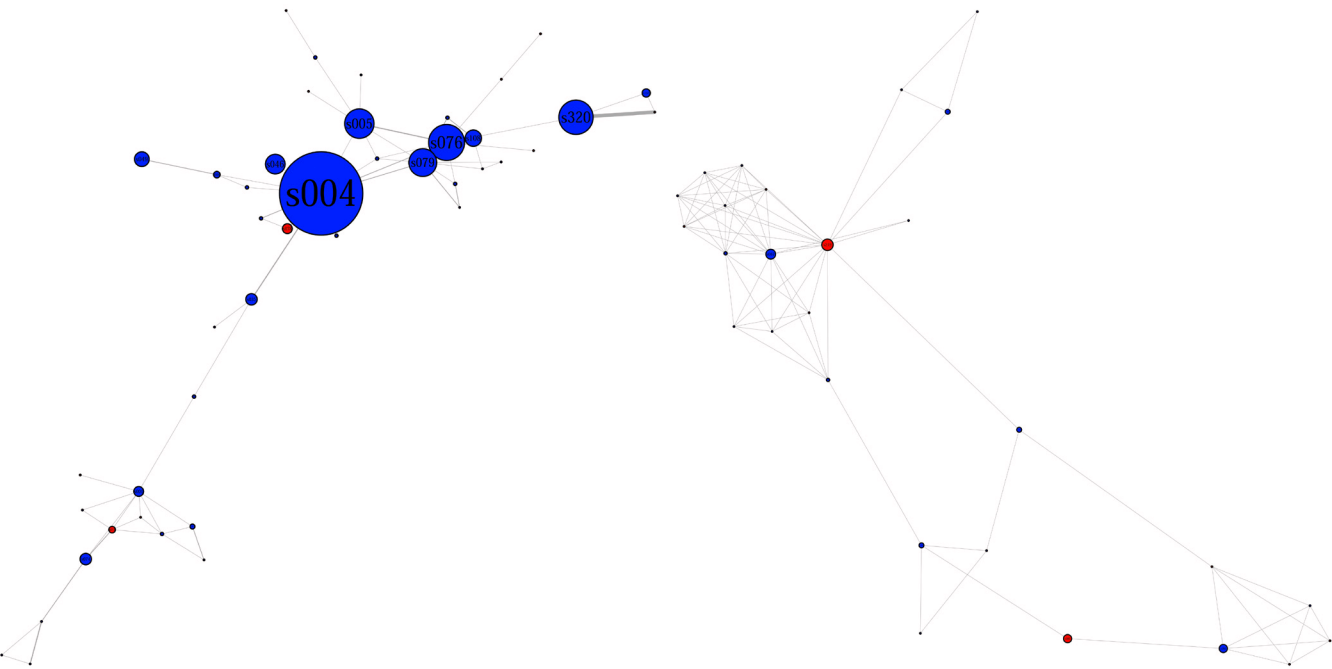


Figure 4. Visualization of Comp43 (left) and Comp27 (right) with node size equals scientific output. This figure is a visualization of the full 43-node and 27-node components in Fruchterman-Reingold layout. Nodes are color-coded based on author gender (blue for male, red for female). Node sizes are based on node degrees. Edges are represented by a line connecting concerned nodes.

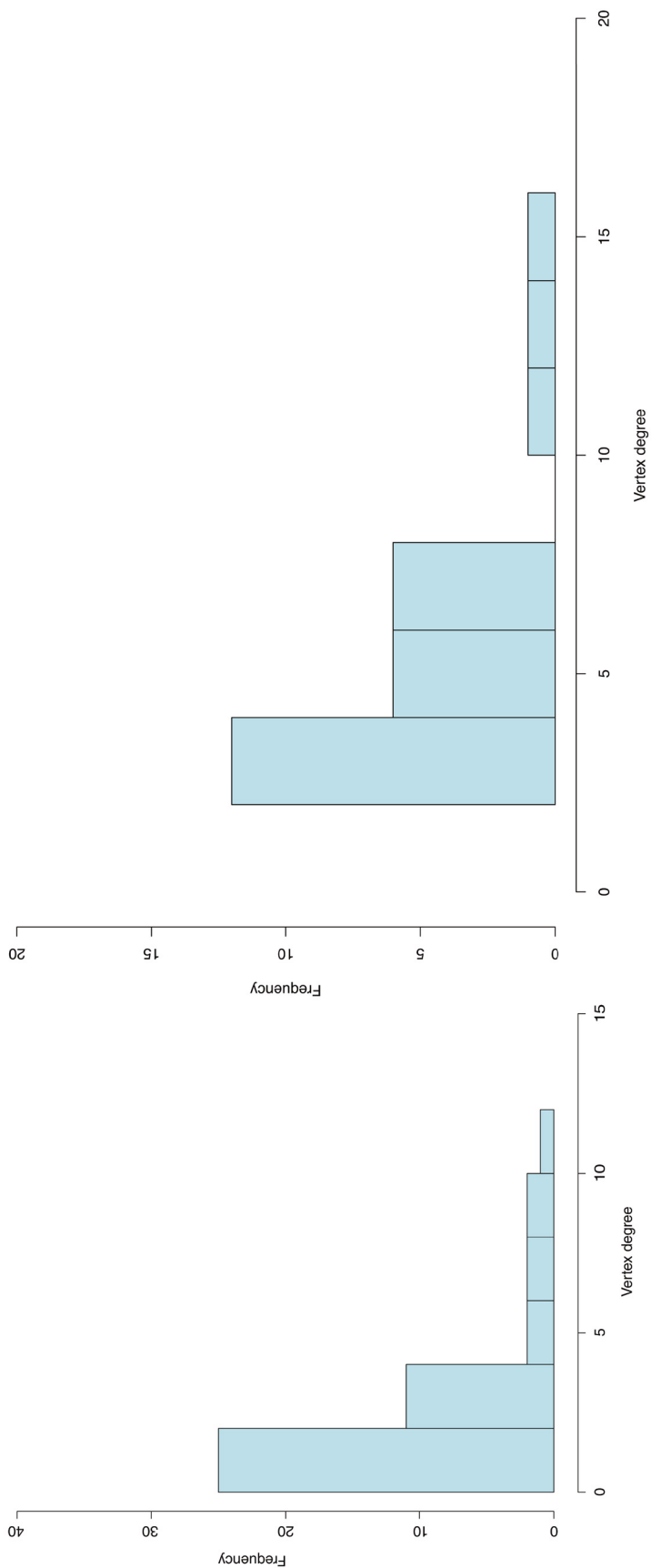


Figure 5. Histogram of degree distribution of Comp43 (left) and Comp27 (right). These are histograms of vertex degree distribution of the 43-node and 27-node components. The degree of each node is measured as the number of co-authored papers, or connections, of each individual in the dataset.

Commands for plotting [Figure 5](#) (left and right) in R can be found in [Supplementary File 8](#) and [Supplementary File 9](#) (“Rcommands_fig5left.doc” and “Rcommands_fig5_right.doc” respectively).

3. Discussion

After performing social network analyses on a sample of 412 social scientists in Vietnam, whose information has been gathered primarily from their Scopus profiles.

First, the study has shown that the network has a low level of connection with only 0.47% of all potential edges realized, and high in clustering with 59% chance a connected triple would close into a triangle. These two characteristics together suggest a reality that the communication and exchange of knowledge and expertise among the Vietnamese social scientists are not very efficient. In addition, the degree distribution reveals that it would be difficult for the network to stay well-connected when a few highly-connected nodes and their edges are removed; or, in network theory’s terminology, the network is not very robust.

Second, in this study, network visualization is shown to be useful not only in facilitating quantitative understanding but also in discovering new insights into the structures of the network. By applying appropriate techniques of graph plotting, the disparity of the level of connections and the structure of the network can be easily visualized. Using the community detection algorithm, an interesting fact about these biggest groups is unraveled: they mostly comprised of smaller and tightly connected communities with one or two vertices connecting these altogether.

Third, close investigations show that the two largest components in the network have different characteristics from the 412-node-graph. Both smaller networks have more connections than the big one, but in terms of clustering, the 43-node-graph has a much higher level of clustering. Despite these differences, all the three networks resemble in low level of robustness and high disparity in terms of degree distribution, which means when the most connected people are removed from the networks, these latter would immediately be decomposed into several smaller groups. Most strikingly, the two smaller networks seem to be led by the most productive researchers in them, who also have the most connections.

Given the mostly high transitivity of all three networks, it could be remarked that the original 412-node network could be considered more or less a sum of smaller communities centered around well-connected nodes. On a more ego-centric and contextual note, there seems to be a relationship between the social status (their position in an institution, for example) of an individual in the network and his or her importance to the network (whether he/she has the most connections or being central to many connections in some ways) as well as his or her scientific output, as suggested by the examples of node s004, s067 and s219. These individuals are few and far between in a network of high disparity in vertex degree, and present a stark contrast with their peers in terms of both connections and productivity. They have the potentials to form a group of intellectual elites.

Finally, there is still much to be learned from both the dataset of 412 social scientists and the network that can be constructed from the raw data. For example, though the study has hinted at the difference in scientific output of two networks (comparison of Comp43 and Comp27 in section 2.3), it is worth considering a more systematic examination of the relationship between a network’s properties and the scientific output of the vertices it contains. Thus, finding out whether a correlation among these variables exists does merit further investigation. Another promising area of research is the exploration of diversity in scientific co-authorship. In this study, node color is coded by gender (section 2.2), but other attributes such as age, region, work, titles, etc. can also be added to the analysis as well.

This paper cannot claim to have exhausted the toolkits that social network analysis could provide. There are still many other aspects of the network worthy of further investigation. How would the network turn out if other dimensions such as weights or durability of the relational data are added to the analysis? How useful are certain aspects of the network in predicting scientific performance? How would this network evolve over time? Not only intellectually stimulating, these important questions are of tremendous practical value for policy-makers and educators, particularly when their decision-making concerns education policies and research organizations. Further investigation in this area of research and on this topic is thus necessary.

4. Materials and methods

4.1. Materials: Original data and the network data set

The data for this study was derived from a dataset on the productivity of Vietnamese scientists in the field of social sciences collected by Vuong & Associates. The investigation, which took place within two months from March to April 2017, was conducted under the license V&A/03/2017, issued on 15 March, 2017.

First, we constructed a file that contains data on all the attributes of each author, called a “Nodes list” ([Dataset 1](#): “20170725_net412_NODES.csv”). The data collection process was monitored regularly to ensure its reliability, including the following steps: first, the research team used sources such as personal and institutional websites of authors, websites of journals where their works were published, Google Scholar, and Scopus database to collect data. Then, to check the accuracy of the information, we compare various online sources where each author’s information can be found; for example, Google scholar versus Scopus, personal websites versus institutional websites. After this process, the research team obtained a complete dataset of 412 scholars’ information, consisting of: (i) age, sex, region; (ii) affiliations; (iii) fields of study; (iv) the number of publications in Scopus, (v) the number of research years since the Master graduation; (vi) the number of researchers they collaborated with; (vii) whether or not they have the title of “Professor/Assoc. Professor”. All of this essentially constitutes the node.

Based on this information, we then construct our relational data, which is called an “Edges list” ([Dataset 2](#): “20170729_net412_LINKS.csv”). We consider two authors as exhibiting a

co-authorship tie when they appear together in a scientific publication. Each time the same two authors appear together in a paper, it is counted toward the “weight” of the tie. The example of an edges list can be seen in the following figure. The data was then processed and analyzed using statistical software R (v3.3.1). Figure 6 shows an example of how relational data is handled in the study. To illustrate, in the first row of the table on the left side, a published paper being co-authored by scientists ID s004, s076 and s079 is recorded into the database first. Then on the right side, co-authorship relations among these three scholars are recorded; and the weight is the count of how many times each pair co-authors.

The data for Comp43 and Comp27 were manually extracted from the full dataset. Nodes lists (Dataset 3 “20170719_comp43_NODES.csv” and Dataset 5 “20170726_comp27_NODES.csv”) and Links lists (Dataset 4 “20170719_comp43_LINKS.csv” and Dataset 6 “20170729_comp27_LINKS.csv”) for Comp43 and Comp27 respectively were constructed by picking relevant edges and nodes from the original lists.

Dataset 1. 20170725_net412_NODES.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174929>

This dataset contains all 412 individuals in the study and their attributes. Each individual is considered a node (vertex) in the network.

Dataset 2. 20170729_net412_LINKS.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174930>

This dataset lists the number of co-written articles between all 412 authors of the network, where relevant. Each collaboration is counted as a link (edge) in the network.

Dataset 3. 20170719_comp43_NODES.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174931>

This dataset contains 43 individuals in the 43-node component and their attributes. Each individual is considered a node (vertex) in the component.

Dataset 4. 20170719_comp43_LINKS.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174932>

This dataset lists the number of co-written articles between the 43 authors of the 43-node component, where relevant. Each collaboration is counted as a link (edge) in the component.

Dataset 5. 20170726_comp27_NODES.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174933>

This dataset contains 27 individuals in the 27-node component and their attributes. Each individual is considered a node (vertex) in the component.

Dataset 6. 20170729_comp27_LINKS.csv

<http://dx.doi.org/10.5256/f1000research.12404.d174934>

This dataset lists the number of co-written articles between the 27 authors of the 27-node component, where relevant. Each collaboration is counted as a link (edge) in the component.

4.2. Methods of Analysis

The method employed in this study was statistical analysis of network data. There were several reasons why we choose this method. First, the prevalence of co-authorship in research efforts among

Title	year	Journal	ID			Weight
Does Economic Inequality Affect the Quality of Life of Older People in Rural Vietnam?	2017	Journal of Happiness Studies	s076 ;s004; s079	s004	s076	2
The Effect of Having Children on Women's Marital Status: Evidence From Vietnam	2017	Journal of Development Studies	s004; s046	s046	s004	1
Does firm privatisation benefit local households? The case of Vietnam	2015	Post-Communist Economies	s005; s004; s079; s080	s079	s004	2
Firm agglomeration and local poverty reduction: evidence from an economy in transition	2016	Asian-Pacific Economic Literature	s005; s004; s076	s004	s080	1
				s005	s004	2

Figure 6. An example of the process of handling relational data. In these figures, a fraction of the construction of Edges lists is shown. The table on the left shows how we record 4 published articles in which 5 Vietnamese scientists coded as s004, s005, s076, s079, s080 take part as co-authors. The table on the right shows every pair that have collaborated at least once among these 5 scientists, as well as the number of collaborations of each pair, which are considered the “weight” of the relation.

Vietnamese scientists as shown in the literature review naturally prompts us to ponder on how the co-authors cooperate and the kinds of interactions that exist among them. Second, as we find out that social network analysis has been applied widely all over the world in the study of scientific collaborations, we expect a match between our interest in characterizing collaboration among Vietnamese social scientists and the technical tools this approach provide. Finally, the help of statistical software allows us to create graphic representation of the network, which supplements all the rigorous numerical analysis with a more intuitive way of understanding interactions among actors in the network.

In this study, we will only focus on a descriptive analysis of our network data. The study is strictly limited to the interactions among Vietnamese scholars only. There are two caveats with regards to the method and the scope of the analysis. First, as the collaborations with foreign scholars are not accounted for in this study, certain interesting features of the networks can be lost. For example, a foreign scholar could cooperate with two Vietnamese scholars, but these Vietnamese scholars might not publish together. Thus, a link is missing. The cumulative effects of this kind of missing links can make the network appear much less connected than it actually is. Second, network analysis is first developed to solve problems in areas such as mathematics, chemistry, electrical circuits, operational research, and computer science before being applied by sociologists in mid-20th Century to study social network, hence, we can expect there are inherent limits to the explanatory power of the technique.

4.3. Network characterizations

In order to understand the visualization of a network, it is important to familiarize oneself with the terminologies of statistical network analysis. Here, we provide an explanation of terms that are relevant for the scope and purpose of this paper. More technical explanations of the terms in this paper can be found in *Statistical analysis of network data with R*¹⁵, and *Social Network Analysis: A Handbook, Second edition*¹⁶.

A graph $G = (V, E)$ is a mathematical structure consisting of a set V of vertices (or nodes) and a set E of edges (or links); elements of E are links between a pair of distinct vertices belongs to set V . When two nodes are connected to each other by an edge, they are said to be adjacent. In this study, a vertex represents a Vietnamese social scientist, which means the total number of vertices is 412. An edge represents a relationship between two distinct Vietnamese social scientists. A concept that connects edge and vertex is degree; a degree of a vertex is the counts of the number of edges incident upon that vertex. For instance, if there are three edges incident upon a vertex, the degree of that vertex is three.

Notice that depending on the attribute of the relationships between two vertices, an edge might or might not have a direction, thus there might be a need to specify the ordering of the pair of vertices in each edge in set E . A directed graph is a graph where each edge in E has an ordering to its vertices; an undirected graph is a graph where an edge needs not to be defined by the ordering in the vertices. In this study, since the relationship among co-authors is considered to be neutral, the graph that shows their relational ties will be undirected.

To understand the structure of a network, two fundamental concepts are clique and component. A clique is a subset of vertices that are fully cohesive, in that, all vertices within this subset are connected by edges. For example, a node is a clique of size one, an edge is a clique of size two, a triangle is a clique of size three, and so on. A component is a subgraph, in which, every vertex can be reached from every other. It is easy to see the different between a clique and a component. In a clique, every two nodes must be connected by an edge or in other words, they must be adjacent; while in a component, every two nodes might or might not be connected by an edge, but they must be somehow connected through a path consisting of a number of other edges and nodes.

Regarding the structure of a network, it is natural to wonder about the level of cohesion of the network: How frequent do the edges appear? How likely do three connected nodes close into a clique size 3? These questions can be answered using the concept of density and global clustering coefficient, also known as transitivity. The density of a graph is the frequency of realized edges relative to potential edges. It can be calculated using the following formula:

$$\text{density} = 2l/[n(n-1)]$$

in which l is the numbers of links (or edges), and n is the number of nodes (or vertices). The clustering coefficient (or transitivity) measures the relative frequency with which connected triples of vertices form triangles:

$$cl_T(G) = 3 \tau\Delta(G) / \tau_3(G)$$

in which $\tau\Delta(G)$ is the number of triangles in the graph G ; and $\tau_3(G)$ the number of subgraphs consist of three vertices connected by two edges, i.e. connected triples.

Armed with understanding of relevant technical concepts, we are able to explore the characteristics of the network of 412 Vietnamese social scientists.

5. Conclusions

With the purpose of understanding the structure and characteristics of the network of 412 Vietnamese social scientists, the study has applied the technique of social network analysis to give a sense of the structure of the network, the level of connection as well as the level of clustering in the network. In the last parts of this paper, we zoomed into the two largest components of the network and compare their relevant characteristics together with the network of the entire sample (in line with the spirit of 17).

Remarks corresponding to each characteristic along with insights into the robustness of the network and the spread of scientific knowledge and expertise in the network have been extracted and discussed. The high clustering of the entire network of 412 Vietnamese social scientists and low density shared by both the original network and its two component networks, seem to be closely related to inefficient dissemination of academic expertise. Both of these in turn lead to modest scientific output, which is at the heart of the perpetual discussions on research capacity in Vietnam. Furthermore, the network, low in robustness, is only held together

by a few well-connected scholars, who seem to also hold significant social positions. This suggests the existence of certain intellectual elites who could perhaps propel Vietnamese scientific output.

Data Availability

Dataset 1: “20170725_net412_NODES.csv” This dataset contains all 412 individuals in the study and their attributes. Each individual is considered a node (vertex) in the network. [10.5256/f1000research.12404.d174929](https://doi.org/10.5256/f1000research.12404.d174929)¹⁸

Dataset 2: “20170729_net412_LINKS.csv” This dataset lists the number of co-written articles between all 412 authors of the network, where relevant. Each collaboration is counted as a link (edge) in the network. [10.5256/f1000research.12404.d174930](https://doi.org/10.5256/f1000research.12404.d174930)¹⁹

Dataset 3: “20170719_comp43_NODES.csv” This dataset contains 43 individuals in the 43-node component and their attributes. Each individual is considered a node (vertex) in the component. [10.5256/f1000research.12404.d174931](https://doi.org/10.5256/f1000research.12404.d174931)²⁰

Dataset 4: “20170719_comp43_LINKS.csv” This dataset lists the number of co-written articles between the 43 authors of the 43-node component, where relevant. Each collaboration is counted as a link (edge) in the component. [10.5256/f1000research.12404.d174932](https://doi.org/10.5256/f1000research.12404.d174932)²¹

Dataset 5: “20170726_comp27_NODES.csv” This dataset contains 27 individuals in the 27-node component and their attributes. Each individual is considered a node (vertex) in the component. [10.5256/f1000research.12404.d174933](https://doi.org/10.5256/f1000research.12404.d174933)²²

Dataset 6: “20170729_comp27_LINKS.csv” This dataset lists the number of co-written articles between the 27 authors of the 27-node component, where relevant. Each collaboration is counted as a link (edge) in the component. [10.5256/f1000research.12404.d174934](https://doi.org/10.5256/f1000research.12404.d174934)²³

Author contributions

Q.H.V., T.M.H. conceived and designed the study; all contributors (T.M.H., H.V.N., T.T.V., Q.M.D., H.H.P., Q.H.V.) contributed equally for the remaining tasks of the research and manuscript preparation, i.e. data set preparation, analysis, interpreting results, and writing and checking the paper.

Competing interests

The authors declare no conflict of interest.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

We would like to thank Vuong & Associates for their research initiative The Network of Vietnamese Social Scientists (NVSS), which enabled the research process and provided the raw data for the study. We particularly thank Dam Thu Ha and Nghiem Phu Kien Cuong for their excellent research assistance.

Supplementary Materials

Supplementary File 1: Commands for plotting [Figure 1](#): “Rcommands_fig1.doc” This file contains the R command used to plot the histogram of vertex degree distribution of the full 412-node network shown in [Figure 1](#).

[Click here to access the data.](#)

Supplementary File 2: Commands for computing network metrics: “Rcommands_metrics.doc” This file contains the commands that compute network metrics.

[Click here to access the data.](#)

Supplementary File 3: Commands for graph plotting: “Rcommands_graph.doc” This file contains the commands that set up the data for graph plotting in R.

[Click here to access the data.](#)

Supplementary File 4: Commands for plotting [Figure 2](#): “Rcommands_fig2.doc” This file contains the R commands used to plot the visualization of the full 412-node network in Fruchterman-Reingold layout shown in [Figure 2](#).

[Click here to access the data.](#)

Supplementary File 5: Commands for plotting [Figure 3](#): “Rcommands_fig3.doc” This file contains the R commands used to run a community detection algorithm on the full 412-network and visually present in on the base of [Figure 3](#).

[Click here to access the data.](#)

Supplementary File 6: Commands for plotting [Figure 4](#) – left: “Rcommands_fig4left.doc” This file contains the R commands used to plot the visualization of the 43-node component shown in [Figure 4](#) (left).

[Click here to access the data.](#)

Supplementary File 7: Commands for plotting **Figure 4** – right: “Rcommands_fig4right.doc” This file contains the R commands used to plot the visualization of the 27-node component shown in **Figure 4** (right).

[Click here to access the data.](#)

Supplementary File 8: Commands for plotting **Figure 5** – left: “Rcommands_fig5left.doc” This file contains the R commands used to plot the histogram of vertex degree distribution of the 43-node component shown in **Figure 5** (left). These commands should be used in continuation with the data set-up from “Rcommands_fig4left.doc”.

[Click here to access the data.](#)

Supplementary File 9: Commands for plotting **Figure 5** – right: “Rcommands_fig5right.doc” This file contains the R commands used to plot the histogram of vertex degree distribution of the 27-node component shown in **Figure 5** (right). These commands should be used in continuation with the data set-up from “Rcommands_fig4right.doc”.

[Click here to access the data.](#)

References

1. Nguyen VT: **Nang suat khoa hoc Viet Nam qua cong bo quoc te 2001 – 2015.** *Vietnam Journal of Science and Technology.* 2016; **2006**(10): 49–54.
[Reference Source](#)
2. Hoang VQ, Dung TT, Napier NK, *et al.*: **Business education in the emerging economy of Vietnam: Twenty years of expectations, illusions and lessons.** In: *Innovation in Business Education in Emerging Markets.* Alon I, Jones V, McIntyre J, Eds; Palgrave Macmillan: NY, New York. 2013; 96–109.
[Publisher Full Text](#)
3. Newman ME: **Scientific collaboration networks. I. Network construction and fundamental results.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2001; **64**(1 Pt 2): 016131.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Moody J: **The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999.** *Am Sociol Rev.* 2004; **69**(2): 213–238.
[Publisher Full Text](#)
5. Ding Y: **Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks.** *J Informetr.* 2011; **5**(1): 187–203.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Abbasi A, Altmann J, Hossain L: **Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures.** *J Informetr.* 2011; **5**(4): 594–607.
[Publisher Full Text](#)
7. Yi Y, Qi W, Wu D: **Are CIVETS the next BRICs? A comparative analysis from scientometrics perspective.** *Scientometrics.* 2013; **94**(2): 615–628.
[Publisher Full Text](#)
8. Yan E, Ding Y, Zhu Q: **Mapping library and information science in China: A coauthorship network analysis.** *Scientometrics.* 2010; **83**(1): 115–131.
[Publisher Full Text](#)
9. Nguyen TV, Pham LT: **Scientific output and its relationship to knowledge economy: an analysis of ASEAN countries.** *Scientometrics.* 2011; **89**(1): 107–117.
[Publisher Full Text](#)
10. Manh HD: **Scientific publications in Vietnam as seen from Scopus during 1996–2013.** *Scientometrics.* 2015; **105**(1): 83–95.
[Publisher Full Text](#)
11. Nguyen TV, Ho-Le TP, Le UV: **International collaboration in scientific research in Vietnam: an analysis of patterns and impact.** *Scientometrics.* 2017; **110**(2): 1035–1051.
[Publisher Full Text](#)
12. Vuong QH, Ho MT, Vuong TT, *et al.*: **Gender, age, research experience, leading role, and academic productivity of Vietnamese researchers in the social sciences and humanities: exploring a 2008–2017 Scopus dataset.** *Eur Sci Ed.* 2017; **43**(3): 51–55.
[Reference Source](#)
13. Van Steen M: **Graph theory and complex networks. An introduction.** 2010, ISBN: 9081540610.
[Reference Source](#)
14. Wu X, Liu Z: **How community structure influences epidemic spread in social networks.** *Physica A: Statistical Mechanics and its Applications.* 2008; **387**(2): 623–630.
[Publisher Full Text](#)
15. Kolaczyk ED, Csárdi G: **Statistical analysis of network data with R.** Springer: NY, New York, 2014; **65**.
[Publisher Full Text](#)
16. Scott J: **Social network analysis: A Handbook.** Second edition. Sage: California, 2000, ISBN: 0761963383.
[Reference Source](#)
17. Vuong QH, Napier NK: **Academic research: the difficulty of being simple and beautiful.** *Eur Sci Ed.* 2017; **43**(2): 32–33.
[Publisher Full Text](#)
18. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 1 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017a.
[Data Source](#)
19. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 2 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017b.
[Data Source](#)
20. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 3 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017c.
[Data Source](#)
21. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 4 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017d.
[Data Source](#)
22. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 5 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017e.
[Data Source](#)
23. Ho TM, Nguyen HV, Vuong TT, *et al.*: **Dataset 6 in: Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008–2017 Scopus data.** *F1000Research.* 2017f.
[Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 29 September 2017

doi:10.5256/f1000research.13433.r25359



Ly Thi Tran 

School of Education, Deakin University, Geelong, Vic, Australia

This interesting and important paper addresses the nature and structure of the network of Vietnamese social scientists who have published in Scopus-indexed journals in the period of 2008–2017. This is a critical issue to Vietnam given the government's recent emphasis on enhancing research capacity and scientific outcomes. The topic of networking and co-authorship is very timely given Ministry of Education and Training's new policies requiring doctoral candidates to publish in Scopus and Web of Science-indexed journals, as a compulsory component of their PhD.

Overall, the paper is nicely written. The literature review provides an effective overview of the status of network statistics analysis and makes reference to the broader international context as well as the Asia region. The aims of the research are clearly articulated. There is a good balance between the presentation and discussion of the data. Even though I do not have statistical expertise and am unable to comment on the detailed analysis, generally I found the key findings of the research logically presented and easy to follow.

A key strength of the paper lies in the way the authors use social network analysis to interpret and explain Vietnamese social scientists' research performance. Based on this analysis framework, interesting insights into the structure of the network, the level of connection as well as the level of clustering in the network have been discussed.

I have five comments below for the authors to consider in enhancing this research project or expanding this research area:

1. More discussion of **the specific implications of this research for improving research capacity and doctoral education** would be very useful in the current context of Vietnam. This information would be welcome by policy makers, leaders of universities and research institutes as well as research team, individual researchers and doctoral candidates. Given the scope of this paper, I hope a follow-up article may focus more on this aspect.
2. A critical issue that should be further explored is what facilitates or inhibits **productive and sustainable** networks among the researchers rather than just an one-off or brief collaboration or co-authorship?
3. The authors mainly focus on using social network analysis to explain research performance but less on predicting research performance and impact. Perhaps a follow-up study can address this issue and may include interviews/survey with the scientists to provide more nuanced

understandings about the topic. An issue of great concern is what characteristics and structure of networks are likely to lead to **not only research productivity but also research impact**.

4. A further issue is whether **the network facilitated at the institutional/organisational level or at the individual researcher level is more sustaining and productive?** In this regard, what might be the incentives or support mechanism from the institution and government (for example, targeted funding for network/partnership development or grant/funding scheme in which capacity for research collaboration or network is one of the assessment criteria) needed?
5. One of the findings of this study is that “the network shows the potential of an intellectual elite composed of well-connected, productive, and socially significant individuals.” What is the implication of this finding for policy makers and education leaders to **facilitate an equitable and inclusive networking and collaboration structure** that supports rather than marginalises less established or ‘non-elite’ Vietnamese researchers including early career researchers, researchers from regional universities and researchers who were not exposed to overseas education.

In sum, I have enjoyed reading this paper and I congratulate the authors for this valuable work! Thank you for the opportunity to read and comment on this paper.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 07 September 2017

doi:10.5256/f1000research.13433.r25518



Donaldine E . Samson

Hawaii Pacific University, Honolulu, HI, USA

Given my knowledge about Vietnam's education system, publishing experiences, I have found the paper's results compelling and cogent. This represents one of the first attempts in Vietnam to study its social sciences research efforts from the community and collaboration perspectives.

The authors intentionally employ network data analysis to explain what cross-section data could hardly tell. Scientific soundness of this research is based on earlier work on social networks and basic network metrics. The authors use a precise and unbiased definition of connectedness based on Newman's study¹. The network visualization tools are not only original in this thread of research, they are especially useful in facilitating understanding of the statistical analysis. The paper is well-structured, striking a balance between data and discussions. The results are clear, easy to understand, and presented logically. Despite the focus on Vietnam social science researchers, the approach can also be productively applied to other researcher populations.

This study of co-authorship patterns can be used to help researchers and research directors understand the structure of research collaboration in social sciences in Vietnam and thus develop procedures, platforms and incentives to increase the robustness of networks and reduce the risk of networks disintegrating with the departure of a highly-connected individual researcher.

The authors identify the potential emergence of an "intellectual elite." More study on the effect of this group and ways to increase the benefits and reduce the risks would be interesting as it is not clear whether this intellectual elite is detrimental or beneficial to the maturation of a robust social science research community in Vietnam.

I sincerely hope that the authors will continue this direction of research in social network analysis and look forward to their new results.

References

1. Newman ME: Scientific collaboration networks. I. Network construction and fundamental results. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2001; **64** (1 Pt 2): 016131 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Online teaching and learning

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 05 September 2017

doi:10.5256/f1000research.13433.r25520



Tuyen Quang Tran

Vietnam National University, University of Economics and Business, Hanoi, Vietnam

As a member of the evaluation council on economic studies for Vietnam's National Foundation for Science and Technology Development (NAFOSTED), I am particularly interested in this research as this is one of the key issues we have been facing in recent years. Its merits can also be seen immediately. Some of my comments follow.

The authors have shown their heroic act in collecting and preparing the unique data sets at individual levels, which has never been done before.

The insights are clear and useful, with potentially practical implications for policy makers like us. I like the fact that they focus on basic measures where and when complexities do not necessarily bring more useful understanding.

All data sets are presented, accompanied by R codes, which have made the replication and reproduction of the results both easy and transparent. Having gone through the paper several times, I am now pleased to approve this work.

A further comment: for more valuable analyses in the future regarding this theme of research, I would suggest the expanding of the current data sets to include Scopus citation data.

I would like to suggest that some more limitation regarding the sources of data and methods. For example, ISI web of science covers less journals than Scopus, or panel data should be better for future research.

I wish the authors every success in their future research attempt.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Discuss this Article

Version 1

Reader Comment 14 Sep 2017

Adriana Amaya, ESPOL, Ecuador

The manuscript employed an interesting approach and it is quite well-developed. Like Vietnam, Ecuador is trying to improve the level of scientific manuscripts, then I consider this could be a good way to do it, keeping the rigorous processes that all scientific works need to accomplish.

Competing Interests: No competing interests were disclosed.

Reader Comment 10 Sep 2017

N.K. Napier, Boise State University, USA

I am pleased to see such topics become more prevalent in Vietnam and beyond. As emerging economies grow, the need for indigenous research and networks of social scientists demand more attention. The authors are raising good questions about research clusters, need for collaboration of various types, and the importance of encouraging rigorous thinking. In the future, it would be fascinating to see replications of studies like this in other countries -- do the models emerging from this research hold in other settings.

Competing Interests: No competing interests were disclosed.
